

Analysis of multibeam SONAR data using ICA and the dissimilarity space

By Iain Rice[†], [‡]Roger Benton, [‡]Les Hart and [†]David Lowe

[†]Aston University, UK. [‡]Thales Underwater Systems Limited, UK.

Abstract

This paper considers the problem of low-dimensional visualisation of very high dimensional information sources for the purpose of situation awareness in the maritime environment. In response to the requirement for human decision support aids to reduce information overload (and specifically, data amenable to inter-point relative similarity measures) appropriate to the below-water maritime domain, we are investigating a preliminary prototype topographic visualisation model. The focus of the current paper is on the mathematical problem of exploiting a relative dissimilarity representation of signals in a visual informatics mapping model, driven by real-world sonar systems. An independent source model is used to analyse the sonar beams from which a simple probabilistic input model to represent uncertainty is mapped to a latent visualisation space where data uncertainty can be accommodated. The use of euclidean and non-euclidean measures are used and the motivation for future use of non-euclidean measures is made. Concepts are illustrated using a simulated 64 beam weak SNR dataset with realistic sonar targets.

1. Introduction

SONAR (sound navigation and ranging) is used extensively in underwater acoustics. The fundamental issues commonly encountered in the SONAR domain include low signal-to-noise ratio, multipath reflections from targets, sea surface and floor and the high volume of data for analysis. Large sonar systems can produce hundreds to thousands of beams worth of data, the display of which can easily result in data overload for sonar operators when presented in conventional lofargram displays. Considering the multi-beam time series signals as a sequence of time-dependent vectors in a high dimensional observation space (where the dimensionality is the number of beams being recorded), and this observation sequence is the result of an unknown latent generative model consisting of multiple signals mixed with noise, the desire is to find a low-dimensional latent space representation of the high dimensional data which preserves structure in the original data which would be useful to an operator. Such ‘topographic’ (structure-preserving) projections need to be generically nonlinear and ideally capable of handling the mapping projection of new data without being ‘re-trained’ for each block of new data. Additionally, since the observation data is always subject to uncertainty, if there is a framework and information to represent uncertainty in the observation data, the projection method should accommodate the uncertainty in the latent space construct. Finally, since each observation vector (sonar beam scan) only has significance in the context of its relative information content compared to the context of a large number of other scans (ie, human operators perceptually ‘integrate’ lofargrams to smooth out noise, spot anomalous behaviour, occluded tracks etc) we are interested in methods which use relative measures of dissimilarity between pairs of observations, rather than assuming the observed values

of a single isolated observation vector are the fundamental entities. In this paper we introduce one such nonlinear topographic visualisation approach which is flexible enough to use different metrics for representing ‘similarity’ between observations vectors, which need not be metric or even positive definite, and can also use measures of ‘distance’ between distribution functions.

2. NeuroScale: Topographic dimensionality reduction

We seek a dimension-reducing, *topographic* transformation of data for the purposes of visualisation and analysis. By ‘topographic’, it is implied that the geometric structure as determined by pairwise relationships of the data is preserved in the transformation. This is a requirement that relative ‘dissimilarities’ are preserved on average, if possible.

The NeuroScale approach employs a nonlinear transformation $\{\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m : \mathbf{f}(\mathbf{x}) = \mathbf{y}\}$ from the original configuration space that maps into the feature space. We choose the class of nonlinear parameterised transformations provided by Radial Basis Function networks. This has the advantage that a *transformation* is obtained, allowing interpolation. The model parameters are adjusted to minimise the global STRESS analogous to the classical multidimensional scaling method, (but now with the advantage that a transformation is provided and not just a mapping):

$$E = \sum_{p=1}^P \sum_{q < p} [d_n(\mathbf{x}_p, \mathbf{x}_q) - d_m(\mathbf{y}_p, \mathbf{y}_q)]^2,$$

where $d_n(\mathbf{x}_p, \mathbf{x}_q)$ is the ‘distance’ between data points in the original space (often, but not necessarily, taken to be a euclidean distance between points $\|\mathbf{x}_p - \mathbf{x}_q\|$) and $\{d_m(\mathbf{y}_p, \mathbf{y}_q) = \|\mathbf{y}_p - \mathbf{y}_q\|\}$ are the distances in the latent visualisation space (which can also be relaxed if prior knowledge dictates otherwise). The visualisation space dimensionality is often taken to be $m = 2$, but $m = 1$ and $m = 3$ dimensions can also be useful. The points \mathbf{y} are generated by the Radial Basis Function network, given the data points as input. That is, $\mathbf{y}_q = \mathbf{f}(\mathbf{x}_q; \boldsymbol{\theta})$, where \mathbf{f} is the nonlinear transformation effected by the Radial Basis Function model with parameters, i.e. weights and any kernel smoothing factors, $\boldsymbol{\theta}$. The (squared) ‘distances’ in the feature space (assuming euclidean discrepancies) may thus be given by

$$\begin{aligned} d_m^2 &= \|\mathbf{f}(\mathbf{y}_q) - \mathbf{f}(\mathbf{y}_p)\|^2 \\ &= \sum_{l=1}^m \left(\sum_k \lambda_{lk} [\phi_k(\|\mathbf{x}_q - \boldsymbol{\mu}_k\|) - \phi_k(\|\mathbf{x}_p - \boldsymbol{\mu}_k\|)] \right)^2. \end{aligned}$$

The topographic nature of the transformation is imposed by the *STRESS* term which attempts to match the inter-point dissimilarities in the latent visualisation space with the dissimilarities in the input space. Note that nowhere is an isolated point vector needed; only measures or pairwise dissimilarity.

Also, note that central to this transformation is the assumption of a provided ‘distance’ functions d_n, d_m in input and latent spaces, and also the choice of interpolating basis spline functions, $\phi(\dots)$.

2.1. Dissimilarity measure

A priori, there is no reason to suppose that the dynamical evolution in the observation or latent spaces should occur on a Euclidean manifold (zero curvature). Prior knowledge on

sensors for example may indicate a geometry of input data space different from euclidean. So using a euclidean measure of distance to characterise dissimilarity between vectors in the input space may be an incorrect simplification. The Bregman divergence is a class of measures which allows for incorporating a degree of ‘manifold curvature’ into measures of similarity. The Bregman divergence between two points \mathbf{p}, \mathbf{q} can be expressed as

$$d_F(\mathbf{p}, \mathbf{q}) = F(\mathbf{p}) - F(\mathbf{q}) - \langle (\mathbf{p} - \mathbf{q}), \nabla F(\mathbf{q}) \rangle \quad (2.1)$$

where $F(\cdot)$ represents a pre-defined function and $\langle \dots \rangle$ represents the inner product. The positivity of Bregman divergence as a measure of ‘distance’ stems from the constraint of the convexity of $F(\cdot)$. The Bregman divergence measures the difference between the manifold value at the point \mathbf{p} with a first order Taylor expansion about the point \mathbf{q} and hence reflects the departure due to the curvature of the manifold. However, note that generally the Bregman divergence is not a metric in that it does not satisfy the triangle inequality and so is not a true ‘distance’. For different choices of convex function, $F(\cdot)$, different measures may be derived, such as the squared Euclidean dissimilarity ($F(\mathbf{X}) = \langle \mathbf{X}, \mathbf{X} \rangle$), KL divergence (if \mathbf{X} is a discrete probability distribution such that $\sum_{i=1}^N X_i = 1$ then choose $F(\mathbf{X}) = \sum_{i=1}^N X_i \log_2 X_i$), Bayes Risk error (choose $F(\mathbf{X}) = -J(\mathbf{X})$ where $J(\mathbf{X})$ is the Bayes risk of the optimal detection rule), and so on.

In an interesting twist, a recent extension to the standard stress function in topographic mappings was given in Wang et.al. (2011) using Bregman divergences with the choice of $F(x) = x \log(x)$ as the convex function measuring the deviations between the dissimilarities in the input space and the discrepancies in the latent visualisation space.

2.2. Non-positive definite splines

It is usual for smoothing splines used in interpolation to be positive definite. The natural domain of analysis of positive kernel smoothing splines is reproducing kernel Hilbert spaces where the vector space is endowed with a positive definite inner product. However in several data analysis situations it is not intrinsic to the problem that the inner product should necessarily be positive. The radial basis function used in the Neuroscale model can deploy non positive definite metrics and basis functions. The natural domain of splines with non-positive definite metrics is a reproducing kernel Krein space [Canu (2005), Hassibi et al (1996)]. This leads to pseudo-Riemannian manifold descriptions of data rather than the archetypal euclidean manifold. In Wang & Lowe (2012) a Minkowski metric was used in a problem where a lorentzian manifold was appropriate for the problem domain. Due to space, this generalisation to the sonar data is the result of future work.

3. Results

3.1. Data

A simulated typically weak SnR 64-beam SONAR dataset has been used for preliminary analysis. The dataset has realistic targets with additive white Gaussian noise. The sampling rate is 4096 Hz. A typical lofargram of a subset of the data is shown in figure 1. There are 3 prominent targets that move through several beams over the full 488 second simulation.

3.2. Time Series Analysis

Although not key to this paper, prior to dissimilarity pattern processing, the raw data was subject to a ‘filtering’ to remove the obvious noise subspace. The method we used followed the approach given in Lowe (1998), James & Lowe (2001), Woon & Lowe

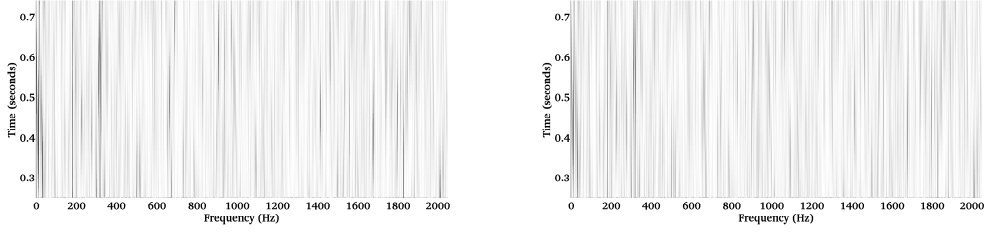


FIGURE 1. Lofargram showing 1 second of the multibeam simulated data, Left subfigure is the original data, Right subfigure shows the reconstructed ‘cleaned’ data observed through the ICA embedding model where prominent target data remains and some background noise activity is removed.

(2004) where we assume a dynamical systems perspective: that the scalar observations are the result of an observation function on a latent higher dimensional state vector evolving on an unknown manifold, but subject to observational and dynamical noise. Using the embedding approach as described in Broomhead & King (1986) to recreate a topologically-equivalent manifold space and then the approach in Lowe (1998), Woon & Lowe (2004), James & Lowe (2001) we extracted a set of common latent sonar ‘sources’ using the single channel ICA approach. The set of sources was automatically split into a set describing the noise subspace, and the remaining set describing the signal subspace. Then the observed sonar signals were reconstructed using only the signal sources. This approach is equivalent to an FIR filtering of the data where the filter coefficients are data-driven. The sources were obtained on a large representative data set and were subsequently held fixed for all future signal processing. The results in this paper were produced using these ICA-subspace filtered signals.

3.3. Exemplar 3D Projection

Each processed temporal slice of 64 time series samples is treated as a dynamical point observation vector in a 64D space which needs to be transformed into a 2D or 3D visualisation space. Each point is regarded as the mean of a gaussian distribution in the observation space, where the (time-dependent and spherical) covariance of the Gaussian is estimated using the variance across all beams as a proxy for uncertainty information. This is merely for illustration of the method being applied to a distribution. Typically, different measures of uncertainty would be available through other means.

Following the discussion in 2, the left-hand subfigure in Figure 2 shows a 3D neuroscale scatterplot visualisation of 1 second of sonar data at full temporal resolution (4096 observations, each of 64 dimensions) using euclidean dissimilarity. The right-hand subfigure in Figure 2 shows the same data segment using a non-euclidean visualisation. The euclidean neuroscale visualisation uses the approach in Wang & Lowe (2012) to map points and their associated uncertainties. The non-euclidean visualisation uses one-sided KL divergences between distributions in input and visualisation spaces and Bregman divergence to replace the STRESS measure. The non-euclidean visualisation of data and its uncertainty is less prone to outliers and also highlights specific ‘interesting’ data points where target data of particular note can be observed. (These points reflect target signatures in the original data which represent major departures from the ‘normality’ of surrounding data and hence would be highlighted to the operator for analysis.)

3.4. Beam Clustering

To highlight the ‘interesting’ beams, we now consider a more traditional use of the dissimilarity representation, following the approach in Duin (2000), Duin & Pekalska

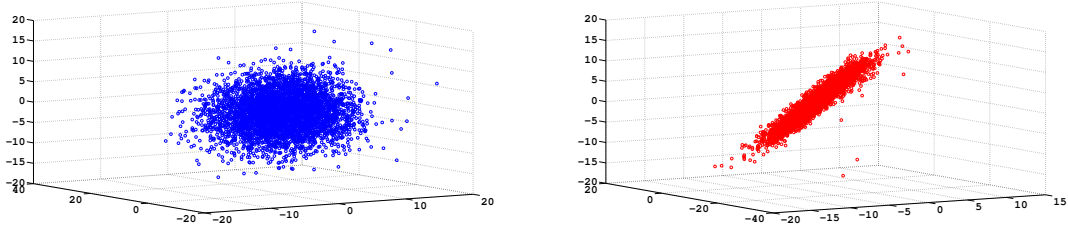


FIGURE 2. Euclidean (left) and non-Euclidean (right) comparisons of the Neuroscale 3D visualisation of 4096, 64D data samples. The Euclidean visualisation is more prone to outliers, and the ‘interesting’ data points are more apparent in the non-Euclidean visualisation.

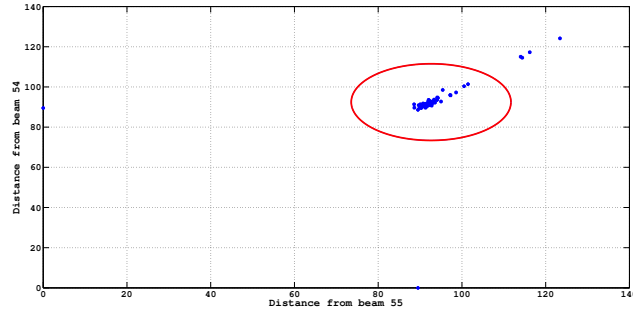


FIGURE 3. 3D PSD dissimilarity of the 64 channels measured relative to three prototypes, channels 62, 55 and 56. Most channels are clustered in one region with small dissimilarity. Points of significantly different dissimilarity represent the actual targets which are located in these channels. The points located on the axes correspond to channels 62, 55 and 56.

(2009) for dissimilarity clustering. For this same segment of data we now consider the power spectrum of each channel of data (after the temporal ICA-processing to remove obvious noise subspaces), and measure the dissimilarity between each channel’s spectrum relative to a set of prototype channels. These prototypes were determined automatically using a modesseek algorithm to select three reference channels (channels 62, 55 and 56 happened to be selected) against which the dissimilarities were measured. This set of 64, 3D points is given in Figure 3. Figure 3 indicates that a cluster of beams is present where many beams are situated within a small range of dissimilarity between the two representatives. A smaller cluster of points with larger dissimilarity is present containing beams 1,2,32 and 33. The 3 targets present in the simulated data are centred at these beams 1 and 2, 32 and 33 and in the two representatives 54 and 55.

4. Conclusions

Taking a dissimilarity view of pattern processing has significant potential for improving the ability of operators to analyse high dimensional and poor quality sonar data, both through providing augmented low dimensional visualisation representations to the conventional lofargram display, and through automated alarm systems to automatically highlight data regions of significant departure from ‘normal’ behaviours. However this requires an investigation into appropriate domain-specific measures of dissimilarity. In

this paper we have introduced a dissimilarity-based visualisation approach which allows the exploration of a range of interesting and nonconventional measures, some of which need not be distances at all, and others that can compare between distributions. This latter ability is important in maritime situations in which data is categorised not just by point samples, but augmented by some measure of uncertainty or importance of a given datapoint.

REFERENCES

- BROOMHEAD, D. S. AND KING, G. P. 1986 Extracting qualitative dynamics from experimental data. *J. Phys. D* **20**, 217–236.
- CANU S, ONG CS, AND XAVIER M, 2005 Splines with Non Positive Kernels, in *5th International ISAAC Congress* World Scientific.
- HASSIBI B, SAYED A.H. AND KAILATH T 1996 Linear estimation in Krein spaces - part I: Theory IEEE TRANSACTIONS ON AUTOMATIC CONTROL, **41**(1) 18–33.
- DUIN, R.P.W. 2000 Classifiers in almost empty spaces. *Proc. 15th Int. Conf. on Pattern Recognition, 2000.* **2**, 1–7.
- DUIN, R.P.W. AND PEKALSKA, E. 2009 Datasets and tools for dissimilarity analysis in pattern recognition. *SIMBAD Project, Information and Communication Technologies Collaborative Project* **3.3**.
- LOWE, D. 1998 Feature space embeddings for extracting structure from single channel wake EEG using RBF networks, *Neural Networks for Signal Processing VIII*, 1998. Proceedings of the 1998 IEEE Signal Processing Society Workshop, 428–437.
- JAMES, C. AND LOWE, D. 2001 Single channel analysis of electromagnetic brain signals through ICA in a dynamical systems framework. *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE* **2**, 1974–1977.
- PEKALSKA, E. AND DUIN, R.P.W. 2005 The Dissimilarity Representation For Pattern Recognition. *World Scientific*.
- PEKALSKA, E., DUIN, R.P.W. AND PACLIK, P. 2006 Prototype Selection for Dissimilarity-based Classifiers. *Pattern Recognition* **39**, **2**, 189–208.
- WANG, X. FYFE C, AND CHEN W. 2011 Applying Bregman Divergences to the Neuroscale Algorithm, *Proceedings of the 11th UK Workshop on Computational Intelligence*, 190–195.
- WANG, X. AND LOWE, D. 2012 Signal Processing Issues of High-Dimensional Visual Informatics: A Study in Protein–DNA Binding Patterns. *9th IMA International Conference on Mathematics in Signal Processing*.
- WOON, W.L., LOWE, D. 2004 Can we learn anything from single-channel unaveraged MEG data? *Neural Computing and Applications*, **13**(4), 360–368.